

First Hit Fwd Refs  

L3: Entry 29 of 174

File: USPT

Apr 8, 2003

DOCUMENT-IDENTIFIER: US 6545699 B2

TITLE: Teleconferencing system, camera controller for a teleconferencing system, and camera control method for a teleconferencing system

Abstract Text (1):

A teleconferencing system has a speaker direction detection means, which detects the direction of a speaker from the phase difference of voice inputs to a plurality of microphones, and an imaging control means, which directs a camera in the detected direction. A movement pixel detection means detects a moving part of the image captured by the camera, using a frame differencing, the distribution of which is measured by a movement distribution measurement means, and the position of person is detected by a speaker position establishing means. Even in the case in which a speaker is not captured at the center of an image, the imaging control means can move the camera so as to enable capturing of the person at the center of the image with an appropriate picture angle, thereby enabling proper capturing of the image of the speaker.

Application Filing Date (1):

20010529

Brief Summary Text (7):

In a method disclosed in the Japanese unexamined patent publication (KOKAI) No. 7-140527, the direction from which a sound is heard is detected from the phase difference in voices input from a plurality of microphones.

Brief Summary Text (13):

Additionally, in the method using the phase difference, there is a large error, and it is difficult to position the speaker in the center of the camera.

Brief Summary Text (20):

Specifically, a first aspect of the present invention is a teleconferencing system having a plurality of sound-collection means, at least one speaker-imaging means, an image-display means, and an imaging control means, which, based on voice direction information of a speaker obtained from the sound-collection means, changes the imaging direction of the imaging means that images the speaker, wherein the imaging control means is controlled so as to direct the imaging direction of the speaker-imaging means toward the direction of a speaker predicted by the sound-collection means, and wherein the imaging control means is configured so that movement pixels are extracted from the captured image, and a distribution of the movement pixels is determined, so as to identify the direction of the speaker within the image, and so that, based on the direction information of the speaker, the speaker is displayed in a prescribed position within the image area.

Brief Summary Text (21):

A second aspect of the present invention is a teleconferencing system having a plurality of sound-collection means, a speaker-imaging means, which images a speaker, a speaker direction detection means, which, based on information from the sound-collection means, predicts the direction of a speaker, a first imaging control means, which, based on information of the speaker direction detection means, changes the facing direction of the speaker-imaging means, an image-display

means, which, in response to a control signal of the first imaging control means, displays a captured image of the speaker-imaging means caused to be faced to a prescribed direction, a movement pixel detection means, which detects movement pixels from the captured image, a movement distribution measurement means, which measures the distribution of movement from the movement pixels detected by the movement pixel detection means, a speaker position establishing means, which, based on the measurement results from the movement distribution measurement means, establishes the position of a speaker in the image, and a second imaging control means, which, based on information of the speaker position establishing means, performs further control of the facing direction of the imaging means.

Brief Summary Text (26):

a fourth step of extracting movement pixel information from the captured image information,

Brief Summary Text (27):

a fifth step of calculating a movement distribution from the extracted movement pixel information,

Brief Summary Text (30):

By adopting the above-noted technical constitution, a teleconferencing system and a camera controller and camera control method for a teleconferencing system according to the present invention has a speaker direction detection means, which detects the direction of a speaker from a phase difference input into each one of a plurality of the microphones or from the voice levels detected by the plurality of microphones, and an imaging control means, which directs a camera to the detected direction, wherein a moving part of a picture picked up by a camera directed at a speaker by means of his or her voice is detected, a movement distribution measurement means measuring the movement distributions thereof in the horizontal and vertical directions, and the position and size of a person being detected by a speaker position establishing means, from the horizontal-direction and vertical-direction movement distribution. Even if the speaker is not captured within the image, the imaging control means can be moved so that the speaker is captured in the center part thereof, with a proper size.

Detailed Description Text (3):

Specifically, FIG. 1 is a block diagram showing the general configuration of an example of a teleconferencing system according to the present invention. This drawing shows a teleconferencing system 100 having a plurality of sound-collection means 1, at least one speaker-imaging means, e.g. a camera, 2, an imaging control means 7, which, based on voice direction information of a speaker obtained from the sound-collection means 1, changes the imaging direction of the speaker-imaging means 2, and an image-display means 8 displaying an image of a speaker. This teleconferencing system 100 is configured so that the imaging control means 7 directs the imaging direction of the speaker-imaging means 2 towards the predicted position of the speaker, predicted by the sound-collection means 1. Additionally, in the teleconferencing system 100, movement image pixels are extracted from the captured image 9 and the distribution of movement pixels thereof is determined, thereby identifying the position of a speaker on the image representation region 12, the position information of the speaker is used as the basis for further control of the imaging control means 7, so that the speaker 11 is displayed at a prescribed position in the image representation region 12.

Detailed Description Text (4):

That is, in terms of the specific configuration, the teleconferencing system 100 has a speaker direction detection means 3, which detects the direction of a speaker from phase differences of voices input to a plurality of sound-collection means, e.g. microphones, 1, and an imaging control means 7, which directs a camera in the detected direction.

Detailed Description Text (5):

This teleconferencing system 100 is additionally provided with a movement pixel detection means 4, which detects a moving part of the image information of the image captured by the camera 2, wherein one or both of the horizontal-direction and vertical-direction distributions of movement are measured by a movement distribution measurement means, and wherein a speaker position establishing means 6, that is, a speaker position establishing means 6 detects the position and size of a person, from the horizontal-direction and vertical-direction movement distributions.

Detailed Description Text (6):

In the present invention, in the case of determining whether the image of a speaker is within the captured image and where it exists therewithin, it is possible, without displaying the captured image on the above-noted image-display means 8, to perform processing by using an arbitrary storage circuit 50, which is either housed within an appropriate image processing circuit provided within the movement pixel detection means 4, or connected to the movement detection means 4.

Detailed Description Text (11):

That is, one example of the configuration of a teleconferencing system 100 according to the present invention, for example, has a plurality of microphones 1 and a camera 2, the imaging angle and picture angle of which can be varied, and further can be a teleconferencing system having a speaker direction detection means 3, which detects the direction of a speaker, using the phase difference between microphones by means of voice data from the plurality of microphones 1, a movement pixel detection means 4, which, from an image captured by the camera 2, detects pixels of a moving object, a movement distribution measurement means 5, which can measure the distribution, in either one or both of the vertical and horizontal directions, of the pixels detected in a moving object, a speaker position establishing means capable of detecting the position and size of a person existing within an image, from the measured movement distribution, and an imaging control means 7, which is capable of performing control so as to direct the camera in the direction of a person as detected by the speaker direction detection means and the speaker position establishing means, and adjust the imaging angle of the camera.

Detailed Description Text (12):

FIG. 2 is a block diagram showing a second example of the present invention. That is, it is possible in a second example of a teleconferencing system 100 according to the present invention, as shown in FIG. 2, rather than using the phase difference of voices as noted above, to provide a speaker microphone detection means 14, which identifies the direction of a speaker as the direction of a microphone, of a plurality of microphones 1 disposed near each participant, which has the maximum voice level input.

Detailed Description Text (15):

The extraction of movement pixels in the displayed image 9, the measurement of the movement distribution, and the operation of correcting the speaker-imaging means 2 based on this measurement are the same as in the earlier noted first example of the present invention.

Detailed Description Text (18):

As noted above, the voice direction information for a speaker can be predicted from the direction of the location of a speaker, using the phase difference between voices of speakers input to each of the sound-collection means 1, and it is also possible to predict that a speaker exists in a direction of a sound-collection means 1 from which the voice level is the maximum value of the plurality of sound-collection means 1.

Detailed Description Text (24):

In a teleconferencing system 100 according to the present invention, although not

illustrated in the drawings, it is preferable that a method by which movement pixels are extracted from a captured image in which the speaker exists, which is captured by the speaker-imaging means 2, and by which a distribution of these movement pixels is determined by a differential image 31 from a plurality of continuous frames after an original image 30 making up the captured image, further determines a distribution image 32 of movement pixels that from the differential image 31, and then performs processing to identify the position 35 of the speaker 40 in the captured image.

Detailed Description Text (25):

That is, it is desirable in a teleconferencing system 100 according to the present invention that the movement pixel detection means 4 generates a differential image 31 between different captured image frames.

Detailed Description Text (26):

More specifically, the differential image 31 is calculated by causing storage of a captured image for each frame in an appropriate storage means 50 connected to the movement pixel detection means 4, and determination of the difference values either between different display image frames selected from a group of display image frames stored in the storage means 50, or between one display image frame stored in the storage means 50 and a display image frame currently displayed on the image-display means 8.

Detailed Description Text (27):

In the present invention, in order to effectively detect movement pixels, at least two display images used when determining the differential image 31 are temporally adjacent, and it is preferable that these are a plurality of display image frames separated by one or several frames, the difference values therebetween being determined.

Detailed Description Text (31):

In the teleconferencing system 100 according to the present invention, in one example of a method for determining the movement pixel distribution, a histogram 33 or 34 of the number of pixels in at least one direction of the vertical and horizontal directions in a captured image is determined, and the position of the speaker 40 within the captured image is determined from the size, density, or shape and the like of the histogram 33 or 34.

Detailed Description Text (33):

A third example of a control apparatus 200 for a speaker imaging means in a teleconferencing system according to the present invention has a plurality of sound-collection means 1, a speaker-imaging means 2, which captures an image of a speaker, a speaker direction detection means 3, which predicts the position of a speaker based on information from the sound-collection means 1, a first imaging control means 7, which changes the facing direction of the speaker-imaging means 2, an image display means 8, which displays an image captured by the speaker-imaging means 2 that is caused to point in a prescribed direction in response to a control signal from the imaging control means 7, a movement pixel detection means 4, which detects movement pixels from the captured image, a movement distribution measurement means 5, which measures a movement distribution from movement pixels detected by the movement pixel detection means 4, a speaker position establishing means 6, which determines the position of a speaker within a captured image, based on the measurement results of the movement distribution measurement means 5, and a second imaging control means 7', which further controls the facing direction of the speaker-imaging means 2, based on information of the speaker position establishing means 6.

Detailed Description Text (37):

The speaker direction detection means can detect the direction from which a sound arrives, that is, the direction of the speaker, from the phase differences between

input voices and the positional relationship between the plurality of microphones.

Detailed Description Text (39):

In the present invention, there is a large error in detecting a person using the phase differences between voices, caused by the direction the speaker is facing and reflections from the walls of the meeting room, making it very difficult to capture the speaker accurately in the center of the image using the speaker direction detection means 1.

Detailed Description Text (41):

That is, a video signal captured by the camera is input to the movement pixel detection means 4.

Detailed Description Text (43):

At the movement pixel detection section 4, the frame differencing is used to extract a movement part of the image.

Detailed Description Text (50):

Although in the above, the immediately previous frame difference (n-1) is used to determine the frame differencing, it is also possible to use an even earlier frame.

Detailed Description Text (51):

In the case of using the immediately previous frame for the difference, because non-extraction of small movements can cause errors, for example, although using a more previous frame requires more storage area to store previous images, this is advantageous in achieving reliable extraction of a moving part.

Detailed Description Text (52):

The movement D (x, y, n) detected by the movement pixel detection means 4 is sent to the movement distribution measurement means 5, at which the vertical-direction and horizontal-direction movement distributions are measured.

Detailed Description Text (64):

Additionally, unless a person is located within the image the processing after the movement pixel detection means 4 cannot correct the camera direction and picture angle, if in addition to requesting the imaging control means 7 to move the camera control is performed of the camera so as to sufficiently widen the picture angle, it is possible to reliably bring the speaker to within the screen. This picture angle can be adjusted in accordance with the amount of picture angle error in the detection of the voice direction.

Detailed Description Text (70):

a fourth step of extracting movement pixel information 31 from the stored image,

Detailed Description Text (71):

a fifth step of calculating a movement distribution 32 from the extracted movement pixel information 31,

Detailed Description Text (74):

In a control method for a speaker-imaging means in a teleconferencing system according to the present invention, it is possible for the first imaging control means 7 and the second imaging control means 7' to be one and the same controller, and based on the voice direction information of the speaker in the first step, to change imaging direction in the second step is determined by phase difference or voice level of the voices input by the plurality of sound-collection means 1 or 13.

Detailed Description Text (75):

It is desirable that the fourth step of extracting movement pixel information in

the present invention can be performed by generating a differential image between different captured image frames.

Detailed Description Text (77):

It is desirable that the fifth step of calculating the movement distribution from movement pixel information extracted in the present invention is performed by forming a histogram with regard to pixel information for either the horizontal direction or the vertical direction, from extracted movement pixel information, comparing this histogram with a pre-established reference histogram, and verifying the position of the speaker within the captured image.

Detailed Description Text (86):

a fourth step of extracting movement pixel information from the captured image information,

Detailed Description Text (87):

a fifth step of calculating a movement distribution from the extracted movement pixel information,

Detailed Description Text (93):

The, at step S6, a judgment is made as to whether or not the imaging direction of the imaging means 2 coincides with the predicted position of the speaker and, if the result is NO, processing of step S5 continues. If the result at step S6 is YES, however, control proceeds to step S7, at which the third step is executed, so as to store an image captured by the speaker-imaging means 2 as a captured image, after which control proceeds to step S8, at which the fourth step is executed, so that the movement pixel information 31 is extracted from the captured image.

Detailed Description Text (94):

After the above, at step S9 a calculation is made of the movement distribution 32 from the extracted movement pixel information 31.

Detailed Description Text (103):

a fourth step of extracting movement pixel information from the stored captured image,

Detailed Description Text (104):

a fifth step of calculating a movement distribution from the extracted movement pixel information,

Detailed Description Text (111):

a fourth step of extracting movement pixel information from the stored image,

Detailed Description Text (112):

a fifth step of calculating a movement distribution from the extracted movement pixel information,

CLAIMS:

1. A teleconferencing system comprising: a plurality of sound-collection means; at least one speaker-imaging means; an image-display means; and an imaging control means, which, based on voice direction information of a speaker obtained from said sound-collection means, changes an imaging direction of said speaker-imaging means imaging the speaker; wherein said imaging control means is configured so as to perform control so as to direct said imaging direction of said speaker-imaging means toward a direction of said speaker predicted by said sound-collection means, and wherein said imaging control means is configured so that movement pixels are extracted from a captured image, and a distribution of the movement pixels is determined by determining an image differencing between a plurality of adjacent frames of said captured image, and determining a movement pixel distribution formed

by movement pixels from a differential image using a histogram of a number of pixels with regard to each of a horizontal direction and a vertical direction in said captured image, and a position of said speaker in an image region is determined from said histogram so that, based on said position information of said speaker, further control is performed of said speaker-imaging control means so that said speaker is displayed at a prescribed position within said image region.

2. A teleconferencing system according to claim 1, wherein voice direction information of said speaker is used to predict a position of said speaker, making use of a phase difference between speaker voices input to said plurality of sound-collection means.

9. A teleconferencing system comprising: a plurality of sound-collection means; at least one speaker-imaging means; an image-display means; and an imaging control means, which, based on voice direction information of a speaker obtained from said sound-collection means, changes an imaging direction of said speaker-imaging means imaging the speaker, wherein said imaging control means is configured so as to perform control so as to direct said imaging direction of said speaker-imaging means toward a direction of said speaker predicted by said sound-collection means, and wherein said imaging control means is configured so that movement pixels are extracted from a captured image, and a distribution of the movement pixels is determined by determining an image differencing between a plurality of adjacent frames of said captured image, and determining a movement pixel distribution formed by movement pixels from a differential image using a histogram of a number of pixels with regard to each of a horizontal direction and a vertical direction in said captured image, and a position of said speaker in said captured image is determined from said histogram, so as to identify said position of the speaker within said captured image, and so that, based on said position information of said speaker, further control is performed of said speaker-imaging means so that a size of said speaker at said position in an image region is adjusted.

11. A teleconferencing system according to claim 9, wherein voice direction information of said speaker is used to predict a position of said speaker, making use of a phase difference between speaker voices input to said plurality of sound-collection means.

18. An imaging means control apparatus in a teleconferencing system comprising: a plurality of sound-collection means; a speaker-imaging means, which captures an image of a speaker; a speaker direction detection means, which predicts a direction of said speaker based on information from said sound-collection means; a first imaging control means, which changes a facing direction of said speaker-imaging means, based on information of said speaker direction detection means; an image display means, which displays an image captured by said speaker-imaging means that is caused to face to a prescribed direction in response to a control signal from said first imaging control means; a movement pixel detection means, which detects movement pixels from said captured image; a movement distribution measurement means, which measures a movement distribution from movement pixels detected by said movement pixel detection means and creates a histogram with regard to pixel information for each of a horizontal direction and a vertical direction, from said detected movement pixel information; a speaker position establishing means, which determines said position of a speaker within said image, based on the measurement results of said movement distribution measurement means; and a second imaging control means, which further controls a facing direction of said speaker-imaging means, based on information of said speaker position establishing means.

20. A control apparatus according to claim 18, wherein said speaker direction detection means predicts a direction of a speaker based on either a phase difference or a voice level of speaker voices input to said plurality of sound-collection means.

21. A control apparatus according to claim 18, wherein said movement pixel detection means creates an image differencing from different captured image frames.

25. An imaging means control apparatus in a teleconferencing system comprising: a plurality of sound-collection means; a speaker-imaging means, which captures an image of a speaker; a speaker direction detection means, which predicts a direction of said speaker based on information from said sound-collection means; a first imaging control means, which changes a facing direction of said speaker-imaging means, based on information of said speaker direction detection means; an image display means, which displays an image captured by said speaker-imaging means that is caused to face to a prescribed direction in response to a control signal from said first imaging control means; a movement pixel detection means, which detects movement pixels from said captured image; a movement distribution measurement means, which measures a movement distribution from movement pixels detected by said movement pixel detection means and creates a histogram with regard to pixel information for each of a horizontal direction and a vertical direction, from said detected movement pixel information; a speaker position establishing means, which determines said position of a speaker within said image, based on the measurement results of said movement distribution measurement means; and a second imaging control means, which further controls said speaker-imaging to change a size of said speaker on said image, based on information of said speaker position establishing means.

27. A control apparatus according to claim 25, wherein said speaker direction detection means predicts a direction of a speaker based on either a phase difference or a voice level of speaker voices input to said plurality of sound-collection means.

28. A control apparatus according to claim 25, wherein said movement pixel detection means creates an image differencing from different captured image frames.

33. An imaging means control method in a teleconferencing system comprising a plurality of sound-collection means, at least one speaker imaging means, and a imaging control means, which changes an imaging direction of said speaker-imaging means so as to capture an image of the speaker, based on voice direction information with regard to the speaker obtained from said sound-collection means, said method comprising: a first step of predicting a direction of a speaker from information collected as voice sound by each of said sound-collection means; a second step of causing an imaging direction axis of said speaker-imaging means to be faced to said predicted direction of said speaker by a first imaging control means driving said speaker-imaging means, based on speaker direction information predicted in said first step; a third step of displaying an image captured by said speaker-imaging means on a display means; a fourth step of extracting movement pixel information from said displayed image information; a fifth step of calculating a movement distribution from said extracted movement pixel information by determining an image differencing between a plurality of adjacent frames of said captured image, and determining a movement pixel distribution formed by movement pixels from a differential image using a histogram of a number of pixels with regard to each of a horizontal direction and a vertical direction in said captured image; a sixth step of determining a position of said speaker within said captured image from said movement distribution information; and a seventh step of a second imaging control means causing said speaker-imaging means to move, based on position information of the speaker within a displayed image, so that said image of said speaker is moved to a prescribed position within said captured image.

34. An imaging means control method in a teleconferencing system comprising a plurality of sound-collection means, at least one speaker imaging means, and a imaging control means, which changes an imaging direction of said speaker-imaging

means so as to capture an image of the speaker, based on voice direction information with regard to the speaker obtained from said sound-collection means, said method comprising: a first step of predicting a direction of a speaker from information collected as voice sound by each of said sound-collection means; a second step of causing an imaging direction axis of said speaker-imaging means to be faced to said predicted direction of said speaker by a first imaging control means driving said speaker-imaging means, based on speaker direction information predicted in said first step; a third step of displaying an image captured by said speaker-imaging means on a display means; a fourth step of extracting movement pixel information from said displayed image information; a fifth step of calculating a movement distribution from said extracted movement pixel information by determining an image differencing between a plurality of adjacent frames of said captured image, and determining a movement pixel distribution formed by movement pixels from a differential image using a histogram of a number of pixels with regard to each of a horizontal direction and a vertical direction in said captured image; a sixth step of determining a position of said speaker within said captured image from said movement distribution information; and a seventh step of a second imaging control means adjusting a zoom mechanism of said speaker-imaging means so as to adjust a size of said speaker in said captured image based on position information of the speaker within a displayed image.

35. A recording medium onto which is stored a program for execution by a computer of an imaging means control method in a teleconferencing system comprising a plurality of sound-collection means, at least one speaker imaging means, and an imaging control means, which changes an imaging direction of said speaker-imaging means so as to capture an image of the speaker, based on voice direction information with regard to the speaker obtained from said sound-collection means, said method comprising: a first step of predicting a direction of a speaker from information collected as voice sound by each of said sound-collection means; a second step of causing an imaging direction axis of said speaker-imaging means to be pointed at said predicted position of said speaker by a first imaging control means driving said speaker-imaging means, based on speaker direction information predicted in said first step; a third step of storing an image captured by said speaker-imaging means; a fourth step of extracting movement pixel information from said stored image; a fifth step of calculating a movement distribution from said extracted movement pixel information by determining an image differencing between a plurality of adjacent frames of said captured image, and determining a movement pixel distribution formed by movement pixels from a differential image using a histogram of a number of pixels with regard to each of a horizontal direction and a vertical direction in said captured image; a sixth step of determining a position of said speaker within said captured image from said movement distribution information; and a seventh step of a second imaging control means causing said speaker-imaging means to move, based on position information of the speaker within a displayed image, so that said image of said speaker is moved to a prescribed position within said captured image.

36. A recording medium onto which is stored a program for execution by a computer of an imaging means control method in a teleconferencing system comprising a plurality of sound-collection means, at least one speaker imaging means, and an imaging control means, which changes an imaging direction of said speaker-imaging means so as to capture an image of the speaker, based on voice direction information with regard to the speaker obtained from said sound-collection means, said method comprising: a first step of predicting a direction of a speaker from information collected as voice sound by each of said sound-collection means; a second step of causing an imaging direction axis of said speaker-imaging means to be pointed at said predicted position of said speaker by a first imaging control means driving said speaker-imaging means, based on speaker direction information predicted in said first step; a third step of storing an image captured by said speaker-imaging means; a fourth step of extracting movement pixel information from said stored image; a fifth step of calculating a movement distribution from said

extracted movement pixel information by determining an image differencing between a plurality of adjacent frames of said captured image, and determining a movement pixel distribution formed by movement pixels from a differential image using a histogram of a number of pixels with regard to each of a horizontal direction and a vertical direction in said captured image; a sixth step of determining a position of said speaker within said captured image from said movement distribution information; and a seventh step of a second imaging control means adjusting a zoom mechanism of said speaker-imaging means so as to adjust a size of said speaker in said captured image.